

AD-A173 653

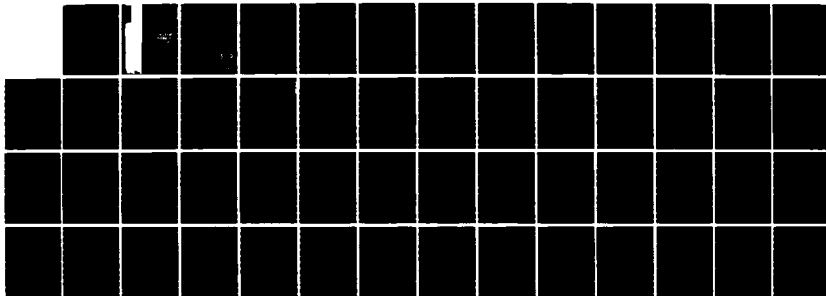
CRITIQUE OF FY 1984 ADVERTISING MIX TEST OF WHARTON
CENTER FOR APPLIED RE. (U) TEXAS UNIV AT AUSTIN CENTER
FOR CYBERNETIC STUDIES A CHARNES ET AL. SEP 86
CC5-RR-546 N00014-86-C-0398

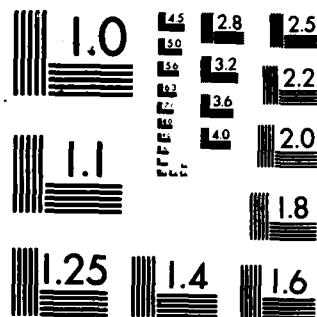
1/1

UNCLASSIFIED

F/G 5/9

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

Research Report CCS 546

CRITIQUE OF FY 1984 ADVERTISING MIX TEST
OF WHARTON CENTER FOR APPLIED RESEARCH

by

A. Charnes
W.W. Cooper
B. Golany
P. Brockett

CENTER FOR CYBERNETIC STUDIES

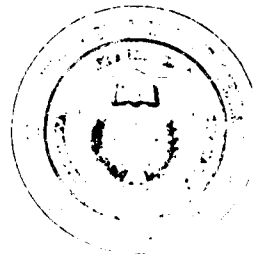
The University of Texas
Austin, Texas 78712

DTIC
ELECTE
30 SEP 1988

S
B

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited



86 10 28 013

Research Report CCS 546

CRITIQUE OF FY 1984 ADVERTISING MIX TEST
OF WHARTON CENTER FOR APPLIED RESEARCH

by

A. Charnes
W.W. Cooper
B. Golany
P. Brockett

July 11, 1986
Revised September 10, 1986

This research was partly supported by ONR Contracts N00014-86-0398 and N00014-82-K-0295 with the Center for Cybernetic Studies, The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the United States government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
College of Business Administration, 5.202
The University of Texas at Austin
Austin, TX 78712-1177
(512) 471-1821

DTIC
ELECTE
OCT 28 1986
S B D

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

CRITIQUE OF FY 1984 ADVERTISING MIX TEST,
JULY, 1986, OF
WHARTON CENTER FOR APPLIED RESEARCH*

1. Introduction

We start our critique by reiterating what we said in our memorandum of April 23, 1986 -- viz., WCAR (Wharton Center for Applied Research) selected the wrong instrument for studying advertising effectiveness in a way that would provide the guidance needed for the very important decisions to which this study is supposedly directed. (See Exhibit II which contains our memorandum of April 23, 1986.) Although not made explicit in the WCAR report, the following assumptions are basic to their effort:

Assumption 1: It is possible to identify advertising as a causal variable and isolate its effects from other variables in the recruitment process in ways that make it possible to measure the effects of advertising alone.

Assumption 2: The instruments used by WCAR are capable of performing these tasks.

Assumption 3: The thus isolated cause-and-effect relations will remain invariant so that their use for effecting decisions in future periods and plans is justified.

Notice, for instance, the "timeless" character of the models used and the way the time invariance assumed in these models enters into the recommendations (page 97 of the WCAR report) to "reduce the advertising working media spending" for FY 1987 by reference to a FY

* As transmitted under cover of memo dated June 13, 1986, to Lt. Colonel J. Simon Office of Deputy Chief of Staff for Personnel which refers to this as a draft final report from Lt. Colonel John A. Ford, Jr. Assistant Director, Advertising Accession Policy, OSD, and states that appendices providing the detailed background were to be mailed at the end of June. See below, however, for a discussion of these appendices which were not received until August.

1984 study. Time dependent variables do not appear in any of the models in the WCAR report. Hence these models are static as well as time invariant. They are therefore misspecified and may be expected to misrepresent any time dependent phenomena that are present such as the "wear out" effects which were prominently referenced in WCAR's earlier literature review. See "The FY 1984 Advertising Mix Test: A Criterion for Evaluating Advertising Policies, September 1984." Philadelphia, The Wharton School, University of Pennsylvania: Wharton Applied Research Center.

Another feature of the budget recommendations in the WCAR report is that no reference is made to any of the other variables (e.g., number of recruiters) that enter into the recruitment process. Thus, the WCAR assumptions lead to recommendations in which supposedly optimal levels and mixes of advertising are to be determined without reference to numbers of recruiters or other resources utilized! This contrasts strongly with the "systems approach" outlined in our April 23, 1986, memorandum as contained in Exhibit II, where it is argued that (a) efficient advertising levels and mixes should be determined as part of the total recruitment effort so that all resources may thus be simultaneously considered and balanced with each other and that (b) this "systems approach" should contain a monitoring element installed in OSD which would make it possible to utilize up-to-date information in order to detect when any service is not utilizing one or more of its mix of resources with full efficiency. ✓

The closest that WCAR comes to pointing toward any of the above assumptions is to be found on page 96 of its report where "adaptive



initiatives" based on further studies of the WCAR type are recommended for the future. Accepting this recommendation, however, will only repeat the present situation where long time lapses will intervene between study initiation and the results needed for current (and future) decisions. In addition to failing to supply currency to the information needed for such decisions, the sampling and other conditions that are needed for the instruments employed by WCAR will fail (e.g., at ADI levels, ADI = Area of Dominant Influences, a term used by Arbitron, Inc. to identify the area covered by a major TV broadcasting station in a marketing area) to mesh with the regular recruitment administrative machinery so that inferences made from the results of such studies will not lend themselves to the fully meaningful evaluation and detailed guidance needed by OSD for the recruitment process.

We wanted to make our basic objections to the WCAR study clear at the outset since the rest of our report is based on accepting the above assumptions in order to evaluate the WCAR study in its own terms. Even on this basis, however, our reading of the WCAR report leads us to the following major conclusions:

- (1) The conclusions with respect to budgetary amounts and proportions in the WCAR report are not supported by the evidence.
- (2) The models and analyses utilized by WCAR are not correct and tend to be misleading.
- (3) Serious questions are present with respect to the data and the design.

Overall our conclusion is that no reliance should be placed on the WCAR report and, in particular, reliance should not be placed

on any of the budgetary recommendations made with respect to advertising in this report.

2. Budget Recommendations

The budgetary recommendations with respect to the military recruitment advertising budget all appear in Chapter VI of the WCAR report. The total amounts and relative proportions between Joint and Service-specific advertising vary from one part of Chapter VI to another. The reasons given for varying recommendations may be found on pages 99 and 100 of the WCAR report but all of them are based on ad hoc and statistically invalid models and analyses.

To avoid confusion we do not attempt to follow each of the varied budget recommendations in Chapter VI of the WCAR report. They are all in the same spirit and point in the same direction and hence we need only focus on the budget recommendations that appear on page 97 of the WCAR report. These recommendations (like all the others) involve drastic reductions in the overall budget and drastic increases in the proportion of the budget which will be used to support Joint rather than Service-specific advertising. In every case the recommendations for these very drastic changes are made with respect only to advertising mixes and magnitudes without mention of other resources (e.g., recruiters).

For such drastic changes in advertising budgets to be recommended without reference to other activities such as sales effort and promotion and incentive programs is unusual, to say the least. Explicit and detailed justification is therefore required. For recommendations like these to be justified on economic grounds one or both of the following properties must be demonstrated by WCAR:

Property 1:

Advertising expenditures are independent of all other resources that might be utilized (e.g., numbers of recruiters) in their effects on recruitment.

Property 2:

Advertising expenditures have a negative or, at best, a zero effect on recruitment responses over wide ranges of advertising expenditures.

In the WCAR report, neither of these properties is demonstrated as being present in the requisite manner. Hence these properties must be considered to represent an additional pair of assumptions and added to the three assumptions set forth on page 1 of our introduction.

The assumption of Property 1 is especially strong and should have been not only tested in the WCAR study but also supported by reference to the marketing literature. No such references are supplied. In fact the equations used by WCAR as a basis for their budgetary recommendations include number of recruiters as an explanation variable and, moreover, number of recruiters and not amount of advertising is the only variable in their model that achieves statistical significance. The assumption of Property 1 is therefore inconsistent with the model utilized by WCAR for its budgetary recommendations.

An exception to the need for joint consideration of other resources can occur when advertising elasticities are negative as noted in Property 2. To arrive at a recommendation to decrease total advertising expenditures by reference to Property 2, however, it is not sufficient for coincidences to occur in which increases in advertising expenditures are associated

with reductions in numbers of recruits. A direct causal relation between these two variables must be established or assumed and reflected in the WCAR models.¹ Even when such causal connections are established, it is not sufficient to say that advertising expenditures have reached a level where diminishing returns are being experienced. For economic justification a matching of advertising expenditures with other resources is required even in the presence of diminishing returns. This matching is not required only when a level of expenditures is reached where further increases in advertising actually begin to actively repel recruits.

A presence of negative elasticities in recruitment advertising over wide ranges of advertising budgets is a highly unusual result. It should at least have been supported by references to the marketing literature. No such references are provided and, indeed, the recruitment models referenced in the earlier (1984) WCAR report [The FY 1984 Advertising Mix Test, A Criterion for Evaluation Advertising Policies, p. 23] leave all of the significant advertising elasticities at positive values.

Note that a negative elasticity means that increased advertising actually repels recruits. Little credence can be assigned to this possibility especially since the discussions in the WCAR report refers only to the scale of advertising expenditures and neither copy content nor type of media employed are considered. The assumption that improved recruitment performance can be achieved by altering the proportions in favor of Joint advertising is also not supported by reference to Joint advertising having either more favorable prices, use of better copy and media vehicles, or evidence of increasing returns to scale.

3. Economic Models and Analyses:

The budget recommendations for advertising in the WCAR report involve drastic alterations in the Joint and Service Specific proportions

¹ See our remarks on page 2 with respect to the static character of the WCAR models.

as well as the total budget. These recommendations supposedly flow from interactions between the two--i.e., interactions between the simultaneous changes proposed in the proportions and in the total budget. We shall shortly direct attention to the statistical treatment that is supposed to provide the evidence for these hypothetical interactions. It is first useful, however, to draw on economic concepts and analyses as we now do in order to show how the WCAR results differ from standard versions of these analyses.

To obtain a simple graphical portrayal of these economic concepts we use the following log-linear model to portray the relation between Contracts (C), amount of Service Specific advertising (S), amount of Joint advertising (J) and number of recruiters (R):

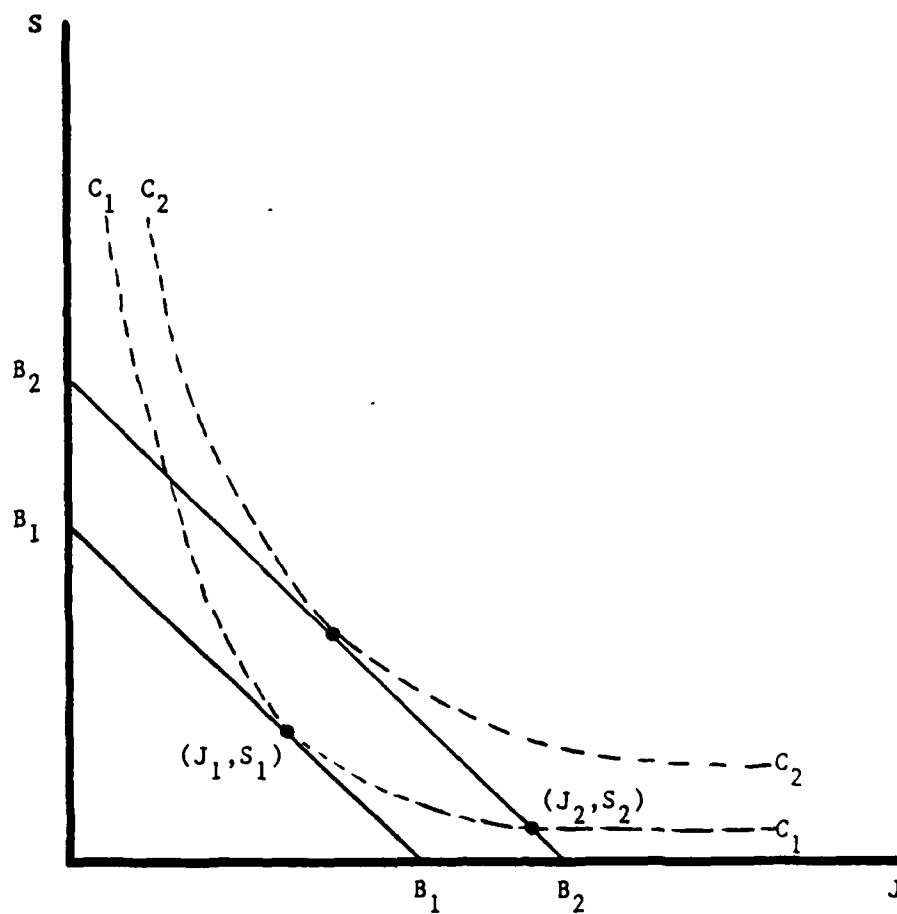
$$(0) \quad C = a \exp (\alpha_1 S + \alpha_2 J + \alpha_3 R)$$

where a is a positive constant, e is the base of the natural logarithms and "exp" denotes the exponential function. $\alpha_1, \alpha_2, \alpha_3$ are also constants, perhaps estimated statistically, but constrained to be non-negative since increases in S , J or R result in increases (and not decreases) in C .

We may use equation (0) above to obtain a "level line," called an "isoquant," by which it is possible to examine all combinations of S , J and R that can be used to produce a given number of contracts C . For constant R the isoquants are of the form shown in Figure 1 where $C_1 < C_2$. The straight lines are $S+J = B_1$ and $S+J = B_2$ for budgets B_1 and B_2 with $B_2 > B_1$. That is, points on these lines represent values of S and J which sum to the budgeted amounts of B_1 or B_2 , respectively.

Figure 1

Contract Isoquants from Equation (0)



To explore all combinations of S and J admitted by B_1 we may start where this line intersects the J axis and move along this same budget line as it moves upward to the left. The highest isoquant attainable on the line for B_1 occurs at the coordinate values (J_1, S_1) . C_1 is therefore the best recruitment level for B_1 . Now suppose the budget is B_2 . If one were at (J_2, S_2) with recruits C_1 it would be possible to move along this line and increase the number of recruits from C_1 to C_2 by moving to the point where the line for B_2 touches the C_2 isoquant curve. Therefore $C_2 > C_1$ is attainable from B_2 .

One could also proceed in the reverse direction and derive the minimum cost of obtaining C_2 at this same point. That is, the same point would yield the cost minimizing split between J and S required to obtain C_2 . To be borne in mind, however, is that the value of R , the number of recruiters, was held fixed arbitrarily and so for overall best results this minimization needs to be expanded by simultaneously considering variations in R with variations in J and S .

In contrast to this we exhibit the WCAR three-piece function as if it were a correct economic formulation in terms of such enlistment contract (or first applicant) isoquants. The formula from WCAR Table 5.24 in the Exhibit I for each piece is of the form

$$(1) \quad Y = a + f(P) + cR$$

Since $P = J / (J + S)$, where J is joint advertising and S is service specific we may write this as

$$(2) \quad f(J/(J+S)) = Y - a - cR$$

or, solving,

$$(3) \quad J/(J+S) = f^{-1}(Y - a - cR) = k$$

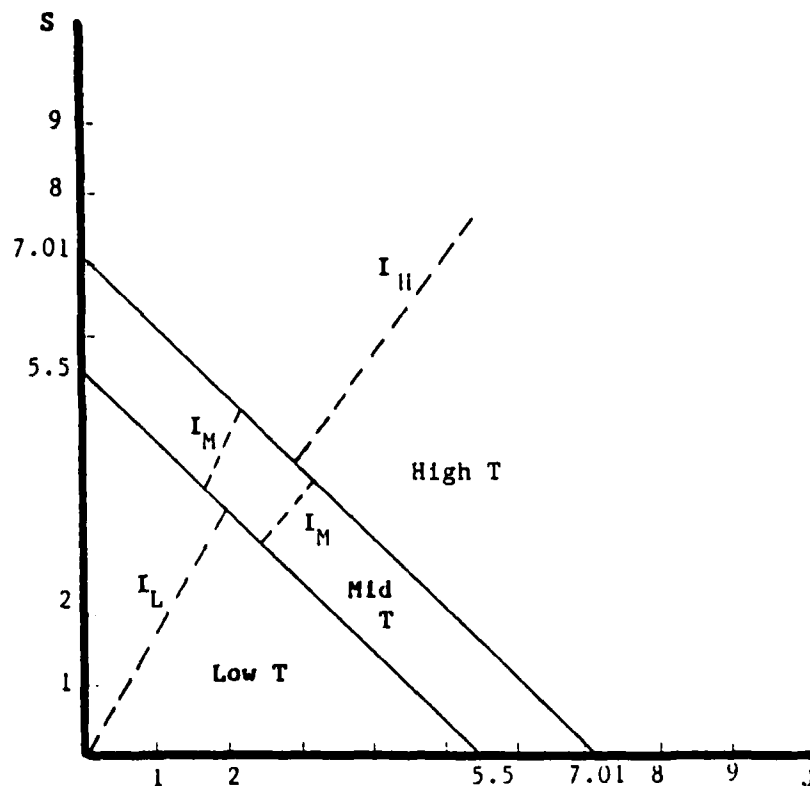
or

$$(4) \quad S = \frac{(1-k)}{k} J$$

Thus for constant R (recruiter man months) and constant Y (first applicants), i.e., on Y-isoquants, each isoquant is comprised of portions of straight lines through the origin! This result holds true whatever the form of the function f (e.g., logarithmic, quadratic, etc.) as long as it depends only on the proportion $P = J/(J+S)$.

As depicted in Figure 2, this means that these WCAR formulae for so-called interaction of Service, Joint and Total Advertising say that in each range the same number Y of first applicants is obtained regardless of how much total advertising is increased or in what proportions (as long as the proportion is constant)! Conversely, these formulae also say that the same number of applicants can be obtained by reducing the total advertising budget to its minimum in the range and finally to zero no matter what the number of applicants! No such bizarre isoquant forms appear anywhere in the economics literature!

These bizarre results and even more curious results follow on recalling that equation (1), hence equation (3), can be valid only

Figure 2WCAR Recruitment Function $T = S + J$ Isoquants (Dashed Lines)

S = Service Specific National Advertising Expenditure

J = Joint National

$$\text{Low } T = \{T \leq 5.5\}$$

$$\text{Mid } T = \{5.5 < T \leq 7.01\}$$

$$\text{High } T = \{T > 7.01\}$$

 I_L = Low T isoquant I_M = Disconnected two piece Mid T isoquant I_H = High T isoquant

for values Y and R such that

$$(5) \quad 0 \leq f^{-1}(Y - a - cR) \leq 1$$

since the value $J/(J + S) = P$, the proportion of Joint Advertising, must be between 0 and 1. Specifically, from WCAR Table 5.24,

$$(5.1) \quad P = \exp [(Y - a - cR)/b] \text{ for } T \leq 5.50 \text{ and } T \geq 7.01$$

which has solutions only for

$$(5.2) \quad Y \leq 1.63 - 2.3R \text{ and } Y \geq .87 + 1.81R \text{ respectively.}$$

For $5.50 < T < 7.01$,

$$(5.3) \quad P = 0.44 \pm 0.57 \sqrt{2.85 + .57R - Y}$$

with solutions only for

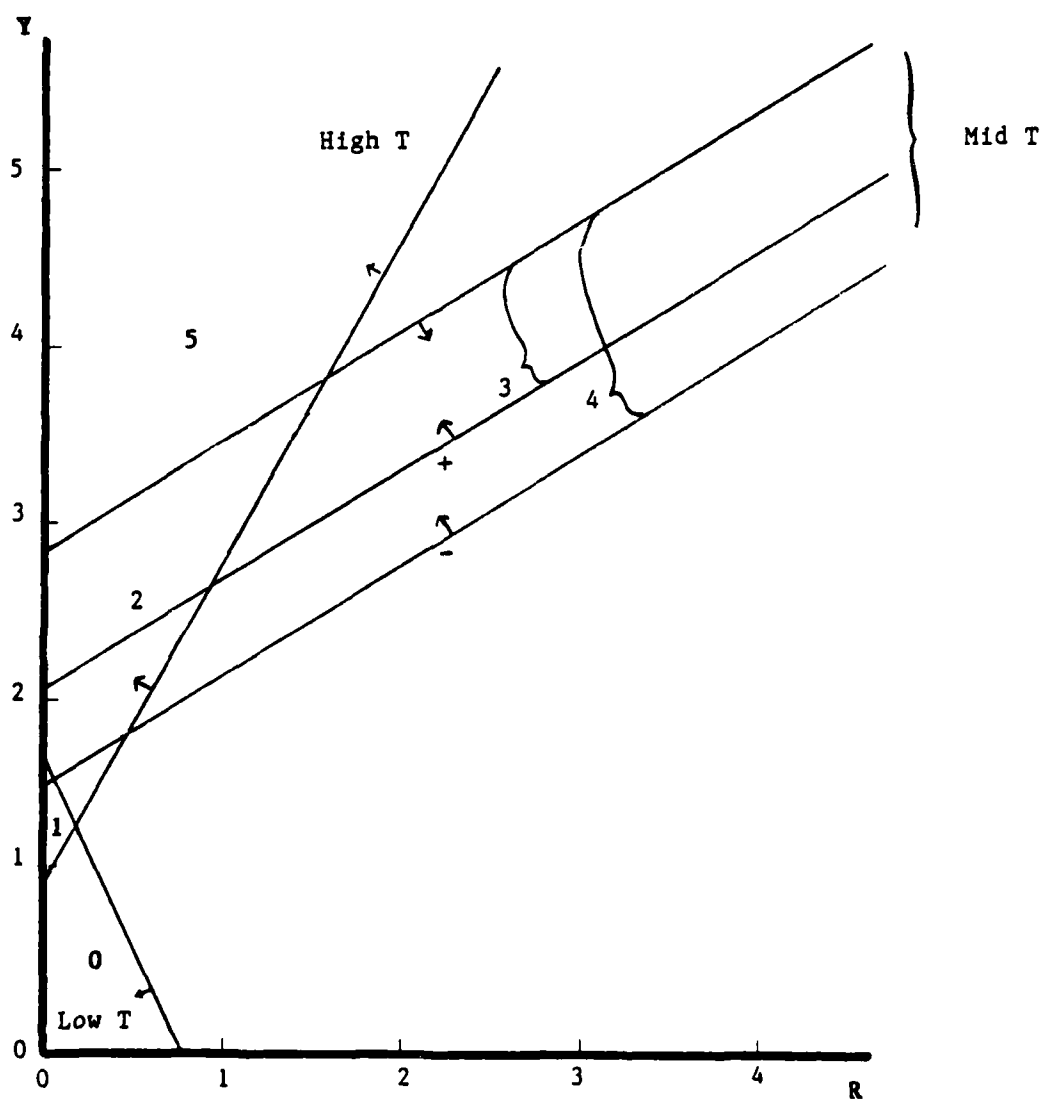
$$(5.4) \quad \begin{matrix} 2.08 \\ 1.51 \end{matrix} \left. \vphantom{\begin{matrix} 2.08 \\ 1.51 \end{matrix}} \right\} + .57R < Y < 2.85 + .57R$$

where 2.08, 1.51 correspond respectively to the $-$, $+$ solutions of (5.3).

These solution regions are depicted in Figure 3 and labelled 0, 1, 2, 3, 4, 5. The arrows on a line specify the side for which the corresponding inequality holds. The regions corresponding to the three Total Advertising ranges are then

Figure 3

(Y,R) Regions of Equation (1) Solutions



<u>Total Advertising</u>	<u>Regions</u>
$T \leq 5.5$	0, 1
$5.5 < T < 7.01$	3 or 4
$T \geq 7.01$	1, 2, 3, 5

For policy recommendations for advertising operations one must apply these to existing first applicants (Y) and recruiter person months (R) situations. From Army recruiter experience many of these occur outside regions 0, 1, 2, 3, 4, 5, where none of these supposed formulae re Total Advertising expenditures can apply at all! And the otherwise faulty and erroneous "mid T" formula can possibly apply only to an extremely narrow range of recruitment performance. WCAR's three part function instrument is totally inadequate for rational consideration of Total and Joint Advertising expenditures.

4. Statistical Models Representations and Analyses

Turning from economics and marketing to statistics, we first note that discussion of the statistical models and analyses in the WCAR report is made unnecessarily difficult by numerous instances of careless writing and editing. For instance, statistical significance is claimed to have been achieved for the Blue cell when discussing the results from the multiplicative (= log linear) model on page 33 of the WCAR report, but the claimed significance is not recorded for the Blue cells in any of the 5 columns to which the discussion supposedly refers in Table 5.3 on page 34 of the report.

Instances of such carelessness are numerous and appear throughout the report. More important than such carelessness perhaps is the repeated failure to report the significance levels (if any) achieved

for the R^2 values. This seems inexcusable since these significance levels are automatically provided in the usual computer printouts, along with associated probabilities of occurrence and other information needed to judge the validity of the statistical relations employed. This information is needed because a failure to achieve statistical significance in R^2 calls into question any use of the relation to which it refers.

The data used in the WCAR study are still not fully available. Even the degrees of freedom associated with the estimates are not given in the WCAR report. Hence, to proceed at all in our discussion, we shall begin as if desired levels of significance have been achieved since otherwise the relations under review are not statistically meaningful and no credence can be given to any inferences made from them.

Additionally, we would be remiss if we did not also record the following two matters: One, the statistical significance levels of 0.05 and 0.10 used in the WCAR report are far too large for a decision as important as the level of advertising needed to support military recruitment. At a minimum, reasons should be explicitly given for effecting these choices for a risk of making the Type I error of rejecting a true hypothesis. Two, no analysis of Type II errors appears in the report, and hence no assessment can be made of the probability of accepting a false hypothesis--e.g., falsely accepting the hypothesis that the Blue cell outperforms all others in all pertinent dimensions. Finally, we record the fact that the significance tests are applied to the parameters

in one-at-a-time fashion even though this procedure is known to yield statistical significance simply because of repeated use of the test. For example, if ten purely random tests are run at the 0.10 level of significance, it can be shown that the expectation is for at least one test to exhibit significance.

Having entered these qualifications into the record, we now proceed to discuss the statistical models and analyses used in the WCAR report. The budget recommendations recorded on page 97 appear to rest on the three-part function represented in Table 5.24 as taken from the WCAR report. See the discussion under "Budgeting Implications" on page 99 of the WCAR report. No justification for the particular functional forms or the ways in which they are combined is supplied by WCAR and so we will check their properties in various ways. To start the analysis of this three-part function, we begin with the piece shown in the lower portion of Table 5.24 after first correcting the typographical errors so this function can be represented as

$$(3) \quad Y = a + bP^2 + cP + dR,$$

where the symbols are defined in the copy of Table 5.24 as taken from the WCAR report and included in Exhibit I.

Apparently the parabolic form shown in (3) was utilized to identify an optimum value at approximately $P = 0.45$ in the second recommendation on page 97 of the WCAR report (whereas the other two pieces of the function are chosen as monotone with optimum solutions only at zero or one).

The expression in (3) is one piece of the three-piece function shown in Table 5.24. The only variables that appear in these expressions are recruiter person months, R, and the proportion, P, of Joint to Total Advertising. The only variable that achieves significance in all 3 pieces of the function is R, recruiter months. Nevertheless, the values assumed for this variable do not appear anywhere in the budget recommendations effected by WCAR on pages 97 ff. On the contrary, only the variable P enters into these recommendations even though none of the estimated values for the coefficients of P achieve statistical significance for (3), above. See the coefficient values tabulated for (3) at the bottom of Table 5.24 in Exhibit I. Note, therefore, that even if $\bar{R}^2 = 0.28$ is assumed to be significant, there is no statistical justification for using these estimated values for the coefficients in (3) to determine an optimal value of P since they are not statistically different from zero.

These same criticisms apply even more strongly to the recommendations on page 100 of the WCAR report which are derived by reference to this same parabolic form but without even supplying the estimated coefficient values on which these recommendations are based. Note that the reference on page 100 is to maximal values of "DoD High School Graduates, and Seniors (HHS), Total Army (applicants) and Army HHS" with different optimal values of P presumably derived from different regression coefficients estimated for the parabolic form (3) in each case. Thus, not only must significance or lack thereof be assessed in the absence of the relevant information, one is also forced to guess what the wholly absent coefficient estimates must have been in each such regression!

The fact that no specific justification is supplied for any of the functional forms used in Table 5.24 has just been noted. The form used at the top of Table 5.24¹, however, is more than just arbitrary. It is illogical and misleading, as can be seen by correcting the typographical errors in the upper part of Table 5.24 and writing this expression as

$$(4) \quad Y = a + b \ln P + cR$$

The logarithmic term in (4) can also be written

$$(5) \quad \ln P = \ln \text{Joint} - \ln (\text{Joint} + \text{Service}).$$

Since P is joint advertising divided by total advertising this part of (4) is illogical in its implication that "Joint" and "Joint plus Service advertising" oppose one another in their effects on recruitment as can be seen in (5). To be noted is that this opposition is a property of the model and not the data. This can be easily seen by using (3) to rewrite (4) as

$$(6) \quad Y = a + b_1 \ln (\text{Joint}) + b_2 \ln (\text{Joint} + \text{Service}) + cR$$

subject to the constraint $b_1 = -b_2$. The Joint and (Joint plus Service). Ad dollars are therefore constrained to work against each other. Finally, stating P in logarithmic form compresses the differences between cells and thus illogically moves in a direction opposite to the Ad-Mix experiment which was justified on the need for having large differences between the cells.

Using percentages in the form of "Joint" divided by "Joint plus Services," as is done in (2), is also misleading in that it biases the results in favor of Joint. To see how this occurs, suppose that the initial ratio is 2/5, i.e., the allocations are \$2 to Joint and

¹See Exhibit I where this Table is reproduced from the WCAR report.

\$3 to Service. Now suppose that an additional dollar is to be allocated. If the dollar is assigned to Joint, this ratio becomes $3/6 = 1/2$ while if it is assigned to Service, the ratio becomes $2/6 = 1/3$. In other words, the allocation to joint increases the ratio from $2/5$ to $1/2$ while the allocation to Services decreases the ratio from $2/5$ to $1/3$. Thus, the model is biased in favor of Joint advertising both in direction and in proportionate effects. Note, in the preceeding example, for instance, that a dollar allocated to Joint advertising produces a 25% decrease and hence has a more than proportionate effect than the 16% decrease that would occur if the same dollar were allocated to Service advertising. To be emphasized again is the fact that these differences in both directional and proportionate magnitude effects are properties of the model and not the data.

Turning next to the break points of $T = 5.5$ and $T = 7.01$ shown in Table 5.24, we have already noted that the choice of these values will be decisive in any cost minimization. We now note that no specific justification for using the particular values shown in Table 5.24 appears in the WCAR report. Vague references are made to results for the multiplicative model (about which more is said later in our report), but no statistical tests or other appropriate validations are applied to these break points. Again, we emphasize that even with the break points, it is the model and not the data which are determining the conclusions. At a minimum the results of several sensitivity analyses using different break points should have been presented. In any case, the WCAR report makes statistical inferences which are really illegitimate because they are based on "non-statistical" choices of break points which were selected after the data were collected.

The statistical tests which yielded the results shown for the coefficient values in Table 5.24 of the WCAR report were obtained from separate applications to each piece of the three-pieced function used in the WCAR model. This is incorrect since the pieces are not separate and independent. In fact, were the study executed according to the design there would be only four points (corresponding to the 4 cells) in Figure 5.1 in the WCAR report and all such ad-hoc analysis would be impossible.

However, the pieces (including break points) are used jointly to arrive at WCAR's budget recommendations, and this is not a statistically valid procedure. To put this in further perspective, we might note that Mr. Carroll stated in the conference at Fort Sheridan¹ that WCAR had tried to extend the parabolic segment in equation (3), above, to the entire range of data as an alternative to the three-pieced function shown in Table 5.24 of the WCAR report, but this attempt was abandoned because it did not achieve statistical significance.

In concluding this section, we might note that the model in Table 5.24 of the WCAR report refers to total applicants and not to non-prior service HSDG I-III A male contracts which is the variable of primary interest to the Army. Hence, the recommendations in the WCAR report are not based on inferences with respect to the class of recruitment responses which form the main concern for the Army. The fact that different media strategies are employed by different services

¹See remarks in Exhibit II contained in the memorandum of April 23, 1986 from A. Charnes to Commander, U.S. Army Recruiting Command.

is not considered in adequate detail and the same is true for the occurrence of fortuitous (but significant) events such as the Marine Corps with its relatively small budget being unable to undertake advertising expenditures in the last two quarters of FY 84. Finally, the possible presence of threshold effects are not even mentioned in the design or execution so that no consideration is given to these effects, which can occur when relatively small budgets, such as the Marine Corps budgets, are reduced to a point where entire classes of media vehicles must be eliminated from possible consideration.

5. Statistical Tests and Estimates

We now turn to more detailed analyses of the models and the tests utilized. This is done seriatim with reference to the corresponding pages in the WCAR report as follows.

1. On page 22 it is stated that the presence of quotas may bias the measure of advertising effects on enlistments. Quotas thus represent a very important variable which is missing from all models. Moreover, no estimates of possible bias are provided when contract estimates are used as the dependent variable in Table 5.3 (page 34) for the multiplicative model and, even more important, this variable (along with others) is omitted from the models used to form the budgetary estimates.

2. On page 23, it is stated that separate adjustments for variations in sample components had to be made. "In particular Cell Green must be adjusted for demographic differences (i.e., race) and adjustments to certain cells must be made to make correct inferences." This was not done for the models of Table 5.24. Hence the policy inferences made from these models cannot be regarded as correct even on the basis of WCAR's own statements.
3. On pages 24 and 25, it is noted that the recruiting process was less productive in 1984 than in 1983. The fact that this occurred in varying rates in different cells indicates that the causal models developed in the design do not include all important explanatory variables and hence are misspecified. Thus credence cannot be assigned to "causal analyses" such as those required to establish the negative advertising elasticities that enter into the budgetary recommendations.
4. On page 31, it is stated that the multiplicative form utilized captures "some nonlinearities as well as possible interactions among the independent variables." Although not stated explicitly in mathematical form in his April 23 discussion at Fort Sheridan, to make the meaning of Table 5.2 in the WCAR report completely clear we write this in the form that Carroll utilizes

$$y = a x_1^{a_1} x_2^{a_2} x_3^{a_3} x_4^{a_4} e^{bz_1} e^{gz_2} e^{rz_3}$$

where

x_1 = recruiters

x_2 = unemployment

x_3 = race

x_4 = urban

e = base of the natural logarithms.

y , the dependent variable, is used to estimate applicants and various classes of recruits in separate regressions.

z_1, z_2, z_3 are dummy variables assigned values of 0 or 1 according to whether blue, green, or red cells are included and the control cell is assumed to be absorbed in the constant, a , to be estimated along with $\alpha_1, \alpha_2, \alpha_3, \alpha_4, b, g$, and r .

The model is estimated by using logarithms which transforms it into the following which is a linear form that has no cross-product interacting terms in its independent variables.

$$Y = A + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + bz_1 + gz_2 + rz_3$$

where

$$Y = \ln y$$

$$A = \ln a$$

$$X_1 = \ln x_1$$

$$X_2 = \ln x_2$$

$$X_3 = \ln x_3.$$

Since ordinary regressions were used to estimate A , α_1 , α_2 , α_3 , α_4 , b , g , and r , and since the above expression is linear, this approach is also best regarded in this fashion--i.e., as a linear relation in which the cross-product interactions between the independent variables are not captured.

5. Employing dummy variables for the actual values for advertising in the cells, as is done in the above model, entails a loss of all information regarding the variability of advertising within each cell. This is a statistically inefficient procedure which runs contrary to the reason for doing the experiment in the first place. The attempts made on p. 29 of the WCAR report to justify this inefficient procedure must be regarded as admissions of faulty design and execution of the study.
6. Much is said about the behavior of the Blue cell as favoring the budgetary recommendations. The behavior of Blue is highly mixed, however, and varies from one table to another and even among parts of the same table. For instance, statistical significance is not achieved for the Blue cell for the highly important category of I-IIIA HSDG applicants in Table 5.2 on page 32. Moreover, for the even more important case of contracts, the Blue cell is not reported as achieving statistical significance in any of the categories in Table 5.3 on page 34.
7. In discussing Table 5.4 the results are again misrepresented (page 33). It is stated that for the high school segment cell Blue is significantly higher and this is not true. In addition the model is worthless since $R^2 = .06$.

Many more such errors and deficiencies can be noted up to and including elementary errors in statistical analysis and inference.

Still more serious are problems like the following:

1. The possible presence of serious problems of collinearity is noted but neither analysis nor remedy is offered to deal with this problem.
2. No evaluation of possible bias in any of the estimates is made and even elementary residual analyses are not supplied. Even when the possible presence of bias is noted, the discussion is inadequate and there is no follow up in subsequent parts of the report. For instance, the biasing effects of the quotas (that are ubiquitously used in the recruitment process) are acknowledged on page 22 of the WCAR report where it is argued that this provides the reason for using applicants rather than the more relevant and important enlistments or contracts for the dependent variable. Nevertheless, the multiplicative model is used without the quota variable to estimate contracts in Table 5.3 on page 34 and without even noting that the latter estimates may be seriously biased by WCAR's own accounting on page 22.
3. Analyses and tests of the validity of the assumptions underlying the statistical tests used (e.g., the t-test) are not covered anywhere in the report. Some of the tests are ad-hoc, nonstandard methods, and some are simply wrong (see, for example, our discussion on "predictive validity" below).
4. Errors of measurement which are present in the independent variables are never discussed even though the presence of such errors calls into question the estimation methods used.

In addition to all of the above, we might note that the report itself is unacceptably low in its professional quality. Not only is it badly written but at many critical points the discussion is so vague that it is virtually impossible to ascertain what was actually done. As a case in point, we might cite the discussion of what is said under the heading of Cross Sectional Pooled Data on page 42. The impression is given that individual regressions were run for each service. If so, this is an inappropriate and incorrect estimating procedure. The dependent variable is the share of an individual service in the total DoD response and these shares must total to 100%. This "summing up" condition is known to present difficulties for satisfactory statistical treatment¹ and the fact that there is no discussion of how this was handled in the WCAR study makes it impossible to evaluate what was done. Since WCAR does not explicitly provide a description of its method for dealing with this summing up one is forced to infer one of the following two possibilities: One, the "summing up" problem was not recognized, which would imply that the analysts were not professionally competent at the levels required for these analyses, or else, Two, the report is written in a way that deliberately obscures what was done in dealing with the summing up problem and this, too, is not acceptable professionally.

6. Predictive Validation Testing:

For added support of its statistical analysis WCAR places heavy reliance on what is referred to as Predictive Validation on pages 89-92 of their report. The criticisms of the statistical analysis made in

¹ See R. Bewley Allocation Analysis (Boston: Ballinger, Inc., 1986).

the preceding sections apply with even greater force to what is done by WCAR in its Predictive Validation.

At the start we note that a model may demonstrate high predictive power in a particular test even when the model is erroneous. This follows from basic propositions in logic wherein true as well as false conclusions may follow from false premises. The purpose of a prediction test is, in any case, to test the predictive power of a model by reference to whether it yields good predictions. As admitted on page 91 of the WCAR report, however, the predictive power is relatively poor since "there is a wide difference between forecast and 1983 cell means."

To analyze what was done by WCAR we confine attention to "point predictions" since that is the only type of prediction dealt with in the WCAR report.¹ Standard methods for conducting such a prediction test would have proceeded from 1983 to 1984 in the following manner. The coefficients of the regression relations would have been estimated from 1983 data. Using these coefficients, the data for 1984 would have been inserted for the variables and used to predict 1984 recruitment results. Exactly the opposite procedure was followed by WCAR because (we are told on page 90), "the very substantial differences between 1983 and 1984 actual responses of all types unrelated to the advertising budget or its allocation." We have underlined this last statement

¹Other types of predictions not dealt with in the WCAR report include the important class of "interval predictions," and others as well. See H. Theil Applied Economic Forecasting (Chicago: Rand McNally, Inc., 1966) and H. Theil Principles of Econometrics (New York: John Wiley, 1971) pp. 134-138.

by WCAR because it is an admission that major variables needed to account for such variations have been omitted from the models and hence causality by reference only to variables in the model cannot be claimed to have been established!

After noting the large differences between actual and forecast values in Table 5.25 on page 92, WCAR then proceeds to employ a series of ad hoc ratio formulations along with a variety of aggregating and averaging devices that obscure these failures. We discuss these devices in further detail as follows.

The ratios employed have no sound basis in the logic of the problem being addressed. To see the illogicality of the ratios employed, we refer to page 91 of the WCAR report where the ratios R_{kj} and S_{kj} are defined. If actual recruits remained constant between 1983 and 1984 the ratio S_{kj} would equal unity. The ratio R_{kj} could be made equal to unity by making an identical prediction for 1983 and 1984 regardless of the numerical magnitudes employed and regardless of the true value. The differences C_{kj} and D_{kj} as defined on page 91 would then be identically zero which WCAR would interpret as a perfect prediction. However the numerical magnitudes of the forecast and the actual recruits could differ greatly. This difference between forecast and actual values would then be concealed in the WCAR approach.

Prediction testing is usually conducted at micro as well as at macro levels in order to insure that the models are properly evaluated for their predictive power in as complete a manner as possible. Instead of using the micro data that were collected to test predictive power either by ADI or by month, WCAR proceeds in an exactly opposite direction.

ADI data are aggregated to test cell levels, monthly data are aggregated to yearly data and service data are aggregated to DoD levels. These data and the ratios formed from them are repeatedly averaged until the reported results come as much from these "smoothing operations" as from the underlying data.

The results of these smoothing operations produce more than obscurity, however. Table 5.26 on page 93, which concludes the portrayal of the predictive tests portrays the results of these repeated uses of averages and ratios. The values in the Table are, however, both meaningless and misleading. For example, values in Table 5.26 are shown as percentages when, in fact, they are differences of ratios of fitted values and differences of ratios of actual values, with no base provided to which any percentage can be referred. Moreover, the stars which are supposed to indicate statistically significant values for the Blue cell are placed on actual numbers which do not involve random variables. Thus, the claimed significance is meaningless. Finally, it is misleading to claim that Cell Blue significantly outperforms the control cell by reference to the sign of Blue compared to other cells, as done at the bottom of page 91 of the WCAR report. These signs refer only to the over- and under-estimation of the corresponding ratio in the control. This has no relevance to the actual recruitment performance in the different cells.

The only information that we can find in this section which is of value for the prediction test is contained in the last two columns of Table 5.25 on page 92 of the WCAR report. As noted earlier, the forecast values for 1983 represented in the next to last column of

Table 5.25 is to be compared to the actual values that appear in the last column. The forecast values are wide of the mark, as noted earlier, and, in addition, strong suggestions of bias are present since the forecast values are lower than the actual values in every case.

7. Data and Design

We are hampered in discussion of data and design issues because we have not yet received the information necessary to complete these tasks from WCAR. Some of the design issues which can be approached conceptually have been assessed in other parts of this report. Other questions that also need to be addressed, however, can only be approached if we have access to the way the design was actually developed and executed in full detail. In particular we need to know what kinds of randomization was used to select and assign the ADIs to cells for the WCAR study.

Consider for instance, the elements in the Blue cell which are compared with the elements that appeared in the earlier study (1979) reported by Carroll et. al. in Marketing Science, vol. 4, no. 4, Fall 1985. Beyond the vague reference to some technical assistance provided by the Rand Corp. (page 9 of the WCAR draft final report), we are not told what randomizing schemes were used for these selections, or, if not random, what influenced the assignment decisions. Now that the technical appendices are finally available we learn that Rand provided two designs but neither was used in the study. See the list of ADIs on page E-26 of the Technical Appendices and note that they differ from both of the design options provided by Rand.

In this regard we wish to point out some matchings in the selection of ADIs for the Blue Cell in the 1984 Ad-Mix experiment and the selection of ADIs for the Reduced advertising Cells in the 1979 Navy Enlistment Marketing Experiment (reported in Marketing Science, Vol. 4, no. 4, Fall 1985 by V. Carroll, A. Rao, H. Lee, A. Shapiro and B. Bayus). We list below the correspondence of ADIs in the two studies.

Thereby, 6 of the 9 ADIs selected for reduced advertising treatment in the Navy 1979 experiment had matching ADIs in the Blue Cell of the 1984 experiment (which is where reduced advertising treatment was applied in 1984). Also, 5 of the 6 others are adjacent to two known high performing ADIs, Portland and Jacksonville.

Similar questions need to be addressed in terms of the way the design was implemented for the other cells. In addition the following further serious questions arise concerning the design itself:

- (1) The cells were not balanced on the major advertising expenditure variables such as total dollars spent on advertising. Thus the cells analyses based on dollars per 17-21 year olds was also not balanced on this critically important variable.
- (2) On page 33 it is stated that "The models have low R^2 values indicating that [there] is not much variation in these ratio data." This is a false interpretation of R^2 . The coefficient of determination R^2 has nothing to do with measuring the amount of variation in the data. Rather it measures the amount of variation which is explained by the model used. In Table 5.4 for I-III A HSDG an R^2 of .06 means that the model explains virtually nothing and is worthless for making any recommendations.

Blue Cell ADIs (1984) vs. Reduced Advertising Cell ADIs (1979)

<u>1984</u>	<u>1979</u>	<u>Correspondence</u>
Dallas-Ft. Worth Tyler (TX) Waco-Temple (TX)	Dallas-Ft. Worth	Tyler and Waco-Temple are adjacent to Dallas- Ft. Worth
Abilene-Sweetwater San Angelo El Paso	Odessa-Midland	All 3 of 1984 adjacent Odessa-Midland
Detroit	Lansing	1984 adjacent to 1979
Harrisburg-York	Johnstown-Altoona Pittsburgh Wilkes-Barre	1984 adjacent to 1979's
	Columbus (OH) Chicago (IL) Phoenix (AZ)	
Bend (OR) Portland (OR) Eugene (OR)		Bend and Eugene adjacent to high performing Portland
Gainesville (FL) Jacksonville (FL) Springfield (MO)		Gainesville adjacent to high performing Jacksonville

- (3) The statistical comparisons of cell pairs is carried to the point of nonsense. Only a comparison of Yellow to Blue is presented. This is incomplete and possibly misleading since there are six comparisons of cell pairs to be made which, if conducted, might lead to opposite conclusions. Nothing is said about the omissions of these additional comparisons. Finally, dummy cell variables were used in an ad hoc, and after the fact, manner since, the WCAR report states (page 36) this was done because of negative elasticities for variables when used as actual per capita expenditures. If the elasticities are known to be positive, then constrained methods of statistical estimation should have been used. Thus even the estimation method used for these ad hoc models is incorrect.

We will next turn to a critique of the technical appendices, which have become available only in August of 1986. Before doing so, however, it is to be observed that there is an open question on how the transitions were effected by WCAR in going from the models that bear the brunt of the statistical causal analyses (which are all of the form represented in section 5, above) and the models that bear the brunt of the policy recommendations (which are all of the form represented in Table 5.24 of Exhibit I). Without a detailed and explicit discussion of how the transition was effected, no justification can be made for using the results of the previous analyses in the models used for the policy recommendations in the WCAR report. Once again the conclusion is the same: no reliance is to be placed on the policy

recommendations contained in the WCAR report.

Finally we turn to the WCAR technical appendices which we comment on in detail in the notation used in the WCAR report as follows:

Comments on the technical appendices in the WCAR Report

Appendix A

Section 1. Test for time trend

The difference between the 1983 and 1984 measures are computed as $d_i = P_{i,84} - P_{i,83}$, where the responses are modeled as binomial random variables with parameters $P_{i,t}$ and the corresponding sample sizes are $N_{i,t}$.

1. This model is wrong and any inferences drawn from it are unjustified. To see that this is true, note that the binomial model presupposes first that the probability of a response is the same for each ADI in the cell in question. This is not true. As recruiter numbers and experience, unemployment, quotas, racial composition etc. vary across the ADIs in a cell, so does the probability of a response within the ADI. This fact has been acknowledged several places in the final report from WCAR. Second, the binomial model assumes that the $N_{i,t}$ responses within a cell are statistically independent (i.e., knowledge of the response values in any one set of ADIs within a cell yields no predictive information about the response values within any other set of ADIs within the cell). This is clearly not true for the advertising mix study data. One can for instance use the data from one subset of ADIs within a cell to better predict that a set of ADIs with High unemployment, many recruiters, and high quotas will have relatively higher recruitment probabilities. This would be impossible if the ADI responses were independent. They

are not independent, and the binomial model is inappropriate.

2. Subsequent inferences from the binomial model are incorrectly drawn. Even if the binomial model were correct (which it is not) the statistical analysis WCAR performs is wrong. Page A-1 alone contains many such statistical and conceptual errors invalidating any and all inference they have drawn from the data. For example on page A-1 it is stated that the variance of $P_{i,t}$ is $P_{i,t}Q_{i,t}/N_{i,t}$. This is true only if the individual ADIs within a cell are independent with the same probability of response. As noted above this is not true. It is also stated that the variance of d_i is the sum of the variance of $P_{i,83}$ and the variance of $P_{i,84}$. Again this is false, since the response probabilities in a particular ADI are highly correlated from one year to the next. In light of correlation, the variances do not simply add, and the inference drawn using this incorrect measure of what is "chance variation in the data" can be highly biased. Finally, on the bottom of page A-1 it is stated that $t(d_i) = d_i/SD(d_i)$ has a student t distribution, and that there is a significant time trend for a particular cell if the absolute value of $t(d_i)$ exceeds the two tailed 90% or 95% critical level. This conclusion is unjustified since there is no theoretical or empirical evidence that correlated binomial random variables divided by an incorrectly computed estimate of the standard deviation has any resemblance to the t distribution at all. Thus all the inferences drawn by WCAR concerning the t-tests must be discarded as worthless. (Incidentally on page 63 of the draft final report it is explicitly stated that "An accurate measure of the effect of advertising on YATS based measures should take into account other important factors...", so the WCAR researchers knew that the exclusion of these other factors could invalidate their whole

analysis, however they still proceeded to use the incorrect and statistically worthless t tests dismissed above.)

Section 2. Pairwise test for difference of differences

Proceeding from the statistically and conceptually flawed binomial model discussed in the previous section, the WCAR report then examines the difference of differences $D_{ij} = d_i - d_j$. In addition to the previous problems with the incorrectness of the binomial model (see comments in section 1.) they now introduce even further errors. The variance of D_{ij} is incorrectly calculated as the sum of the individual variances (which were themselves incorrectly calculated as noted in section 1.). In order for their calculation to be valid, the individual variances for d_i and d_j must be correct, and d_i and d_j must be statistically independent. Neither of these two conditions are met. The incorrectness of the variance calculation was discussed in section 1., and the dependence of the d_i and d_j follows their common dependence on other variables such as numbers of recruiters, unemployment rates, etc. Thus the variance calculation for D_{ij} ignores the important covariance term and consequently is wrong. Since this forms the basis for all of WCAR's inference about significance of the relationships over time, this inference is unsubstantiated and cannot be relied upon. Again, as in the first section, the t-test is unjustified in its usage and all inference based upon it must be discarded as well.

Section 3. Simultaneous tests for difference of differences

1. This entire section is incorrect and contradicts the very assumptions explicitly stated in the reference they quote to justify

their method. The simultaneous difference of differences test is based upon a studentized range test and the reference to Kendall (1968 The Advanced Theory of Statistics, Volume 3, pp. 46-47) is made to justify the test. According to their own reference, the following are the requisite assumptions for deriving the test applied by WCAR:

1. Each of the responses must come from a normally distributed (bell shaped) curve. This is obviously violated by WCAR's data since every response is either a 0 or a 1 and not normally distributed.
2. The observations must all be independent. This is again obviously not true of the WCAR data, as discussed in section 1 and 2 of this comment set.
3. The sample size (number of ADIs in this case) in each group (cell in this case) must be equal. Without the same number of ADIs in each cell, the technique is not justified.
4. The variance in each group (cell) must be the same. This is not true for the WCAR data as they explicitly note on pages A-1 and A-2. Without equal variances, the analysis does not follow.
5. The estimated variance does not depend upon the mean for each group. In the WCAR data the variance for the i^{th} cell is a function of $P_{i,t} Q_{i,t} / N_{it}$ and obviously depends on the mean $P_{i,t}$. Moreover it is easily seen from this formula for the variance that the variances are unequal as well.

Thus in every single respect, the assumptions underlying the studentized range test are violated by the WCAR analysis. No credence can be given to the results of the studentized range test as a consequence of the WCAR group's failure to use a correct analysis procedure.

Additional comments about the irrational usage of the studentized range test are (1) the variances GVAR are again wrongly computed by ignoring important covariation (dependencies) between different cells at the same time and the same cell at different times, and (2) the cell "White" and cell "Yellow" are both included in the analysis (which assumes independence) in spite of the fact that cell Yellow is a subset of cell White and cannot possibly be independent of it. The irrational results which can be obtained by using inappropriate statistical methods is illustrated by the table of studentized range test values (page A-4) where it is found that the algebraic signs of the statistic values change across dependent measures even for the same cell pair. For example, one finds for the Blue-Red comparison that for the dependent variable "1. unaided mention of joining the military" a "significant" difference of -12.1 occurs, while for the variable "3. composite likelihood of joining the military" a "significant" value of +12.1 occurs. This is in spite of the two variables being positively related in theory. In fact in all cell pair comparisons the algebraic sign of the two variables 1 and 3 are opposite. This sort of crazy result follows from the use of inappropriate and unjustified statistical models.

Appendix B

This appendix consists of two tables giving the original and modified cell means. But the statistical design as planned was not carried out. Two ADIs were modified after the data was collected. The Harrisburg-York ADI was post experiment moved from the Blue cell to the Yellow cell, while the Grand Junction ADI which was originally in cell Red was removed from analysis altogether. The Green cell was unaffected by the modifications,

however upon examining the means given in the two tables it is readily seen that the means for the Green cell have been changed from table B-1 to B-2. This is mathematically impossible and shows that the numerical computations exhibited in the report might very well be completely undependable.

Appendix C

This appendix supposedly supports the conclusion that an advertising mix with a higher proportion of joint advertising and a relatively lower absolute total dollar expenditure per capita is optimal. Additionally this appendix attempts to justify the completely ad-hoc method used in the final draft report for considering the interaction of total expenditure and mix of advertising. In both regards the appendix fails miserably.

The statistical methods used (ad-hoc dividing of the data in collections of ADIs cross indexed by total advertising and percentage joint advertising and then using dummy variables to indicated membership in a particular cross classified category) is tremendously wasteful of the information that cost substantial dollars to collect. Moreover, it is (yet again) the wrong technique to use for the goals of the analysis. Why WCAR divided the two dimensional chart under examination into 9 sub-cells sometimes and 6 sub-cells at other times is never explained, and can only be conjectured. The original 72 markets in the experiment were available, and there was no reason to distort the results by grouping the data into non-ordinal cells and then use a statistical method whose sampling properties depend upon normality. By the very act of grouping, WCAR has (yet again) guaranteed that the underlying assumptions of the statistical analysis are not met.

The verbal discussion of the results is a blatant distortion

of the results which can only be judged as purposeful. For example in the second paragraph of page C-1 it is stated that "To demonstrate that the conclusions are not sensitive to the demarkation of cells, two 9 sub-division and two 6 sub-divisions are constructed as shown." First of all the 6 and 9 subdivisions are not comparable. Moreover, variables that were not significant in one cell break-down suddenly become significant in another breakdown. For the applicants of primary interest (upper mental group) and the 9-cell breakdown used in the final report, it is observed that for regression 1-A none of the dummy variables are significant for either applicants or contracts. For regressions 1-B for this same group cell D₇ was significant for both applicants and contracts, but the sign of the significant coefficient was just the opposite between applicants and contracts. This implies the illogical result that applicants and contracts act oppositely. Moreover, upon looking at the data set 1-B it can be noticed that the "significant" dummy variable corresponds to high total budget and low joint service advertising percentage rather than the low total high joint budget recommended in the conclusion section of the final report. However the validity of even this result is in questions since the graphical representation shows only one ADI in the cell D₇. How statistical analysis was performed in this case is unknown.

Another example of the misrepresentation of the analytic result is the statement on page C-1 that "It is observed that the dummy variable corresponding to cells with low total advertising and higher percentage joint advertising are significant." This is false in every important sub-group of applicants, and every important subgroup of contracts,

and for every regression run. Never is it true that the dummy variable corresponding to low total advertising and higher percentage joint advertising (cell D_3 in the 9-cell sub-divisions, the constant term in the 6 cell sub-division 2A, and cell D_3 in the 6 cell sub-division 2B) is significant for the most important high mental group.

8. Conclusion

Data uses as well as data quality issues are present that also need to be addressed. Our understanding is that WCAR claims that it did not use very much of the data that were supplied to WCAR by the Services. Reasons for such non-use appear to turn on questions of administrative expediency and economy from WCAR's standpoint. These are not satisfactory reasons from a technical standpoint. Access to the non-used as well as the used data is needed so that analyses can be undertaken of possible alterations in study results that might emerge from different patterns of data that could have been used.

From a technical standpoint there should also have been a discussion of checks undertaken to see if study conclusions would be affected by using different parts of the data. No such discussion is provided in any part of the WCAR report and hence data based on conclusions in that report cannot be relied upon with any confidence until this is done.

Having imposed on the services and the Government the expenses and burdens involved in submitting these data to WCAR and its subcontractors, we believe there is no good reason why WCAR should not be held responsible for returning these data to the Government in usable form. In fact,

such preparation for data use by others, as well as WCAR, should have been a part of any reasonably good plan for orderly data collection and processing.

For our part we stand ready to undertake the task of analyzing and evaluating the quality of the data used as well as the data not used by WCAR. This cannot be done at the current juncture. In any case we shall have to rely on USAREC to take the initiative while, of course, we will be glad to respond to their leadership in providing the needed ingredients for these parts of a planned review of the WCAR Ad-Mix Study.

Exhibit I

Table 5.24

ANNUAL CROSS SECTIONAL MODELS OF DOD FIRST APPLICANTS
AT DIFFERENT TOTAL ADVERTISING LEVELS

Table 5.24

(Source: WCAR Final Report, July, 1986)

**ANNUAL CROSS SECTIONAL MODELS OF DOD FIRST APPLICANTS
AT DIFFERENT TOTAL ADVERTISING LEVELS**

$$Y = a + b(\ln)P + cR$$

	a	b(ln)P	cR	\bar{R}^2
Low Spending $T \leq 5.50$	1.63	.37*	2.30**	.40
High Spending $T \geq 7.01$.87	-.51**	1.81**	.58

T = Total working media expenditures

Y = Total DoD first applicants

P = Proportion Joint advertising of total advertising

R = Recruiter person months

$$Y = a + bP^2 + cP + dR$$

	a	bP^2	cP	dR	\bar{R}^2
Medium Spending $5.50 < T < 7.01$	2.26	-3.03	2.67	.57**	.28

* - Statistic significant at 0.10 level.

** - Statistic significant at 0.05 level.

Exhibit II

Memoranda to
Commander
US Army Recruiting Command
of
April 23, 1986
and
July 11, 1986

TO: Commander
US Army Recruiting Command

FROM: A. Charnes, Director *A Charnes*
Center for Cybernetic Studies
University of Texas

DATE: 23 Apr 86

SUBJECT: Ad Mix Study of Wharton Applied Research Center

Professor Cooper and I together with Drs P.L. Bruckett and B. Golany have reviewed the material dated March 1986 entitled DOD Advertising Mix Field Experiment, Briefing on the Research Findings by Mr. Vincent Carroll, Principal Investigator, Wharton Applied Research Center (WARC). We also spent April 22, 1986 reviewing these and other findings with Mr. Carroll at a meeting at USAREC Headquarters, Fort Sheridan, Illinois.

As you know, we have also been conducting studies of Advertising and Related Activities bearing on Army recruitment which we made available to all parties some time ago even though we began almost a year later than the studies undertaken by Mr. Carroll and his associates. To date we have still not been able to obtain access to the data that supposedly supports the results and conclusions reported in these briefings as presented by Mr. Carroll. Hence we can only proceed on the basis of these materials, which are evidently part of a slide presentation rather than a detailed report.

We have entered the above remarks because it is our opinion that WARC selected the wrong instrument for studying advertising effectiveness in a way that will provide guidance needed for the very important decisions to which this study is supposedly directed. The length of time needed to conduct this study (which is still not done) should make this abundantly clear. As a result, 1988 and later budget decisions concerning the mix and level of advertising conducted by DOD and the services must be decided on the basis of 1982-1984 data.

There is only one way in which this kind of time lag between data collection and decisions can be justified. Mr. Carroll and his associates will have to claim (1) that there are a fixed set of causes relating advertising to recruitment and (2) that they have truly identified these underlying causal relations in both form and magnitude in a way that must remain invariant over the time intervals that are pertinent to this decision. By contrast our own approach is designed to provide an up-to-date basis for monitoring and evaluating recruitment activities as they are conducted in their regular manner, identify inefficiencies if any are present, and indicate what needs to be done to correct them. The WARC Study provides no details and hence is faulty in this dimension, too, since their approach assumes that all observations were generated from relatively homogeneous behavior in all recruitment organizations and in all activities within the cells covered by their design. Witness, for instance, the failure to give any attention to the possible presence of inefficiencies in operations and how they might be identified in the observations to be collected, either in the original design or subsequent analysis.

To avoid the danger of being misleading with the accompanying risks of erroneous decisions, the WARC Study results must remain invariant from the period for which the data were collected and remain so up through the time when the decisions are in effect. The WARC relations must also hold for all bodies of comparable data. In particular, these relationships must hold for each of the services if they are to provide correct guidance for service responses to the joint advertising activities being recommended by DOD and WARC. Our own analysis as well as our examination of the WARC Briefing Slides raises serious questions as to whether this was accomplished.

In the slide entitled "Why Did We Need an Experiment, Mr. Carroll notes that they could not use historical data on DOD advertising because it does not assure:

- a. Independence: Advertising and recruiters frequently vary together.
- b. Variance: Advertising expenditures frequently vary in only a narrow range.
- c. Measurement: Other relevant data (recruiters, quotas, etc.) are often not available on the same unit of observation.

The subsequent slides then describe the experimental design used to obtain what was then thought necessary to establish the wanted causal relations.

We have discovered shortcomings in the design utilized and its implementation. Here we want to observe that whatever design is to be used its details must be carried out scrupulously to ensure that the causal relations between advertising and recruitment are to be correctly identified when present. In fact, in the course of his April 22 briefing Mr. Carroll admitted that the design was not implemented as intended. We doubt that it is possible to secure conformance with such design conditions by large organizations like the military services which must be guided by other (sometimes urgent) considerations. In this respect, too, the WARC Group selected the wrong instrument for use in such a context.

Among other difficulties we note that the balancing and blocking conditions were not fulfilled as required by the design. We also note that their observations do not have the necessary freedom from error in the independent variables. Hence both mis-specifications and biases are likely to be present in their models and their estimates.

In summary, the instrument used was not appropriate for use in the context in which it is applied and its results are not to be trusted for resolving the policy issues for which they are intended.

There are also technical flaws in the analysis and a variety of additional questions some of which are listed as follows:

"Moving from Results to Conclusions." We comment on this set of slides as follows:

1. "Hypothesis 1: The observed responses are a function only of the advertising expenditure levels."

a. The model is not correctly formulated to test this hypothesis since the expression used makes the response levels depend on the number of recruiters as well as the total advertising expenditures. Furthermore, the number of recruiters and not the advertising expenditures are statistically significant. Hence, as Mr. Carroll admitted in the briefing no conclusions can be drawn as to the level of advertising expenditure despite the way the statement appears in Mr. Carroll's recommendations from hypothesis 1.

b. The formulation of Carroll's model is: $Y = a + b \ln T + cR$, where Y = Total DOD applicants, T = Total DOD Advertising and R = number of DOD Recruiters, all stated per number of 17-21 year old males. The optimal response with their model will be obtained at $T = 0$, if b is negative, or at $T = \infty$ if b is positive. No other result is possible with this kind of model as used in this and the following model. The data cannot affect this in any case.

c. The response variable Y is stated in terms of number of applicants in this area in the following models. This is the wrong measure since the variable of interest is number of graduate, senior male I-III A contracts.

d. All of the models for this and the other hypothesis are faulty in that they do not contain variables that can separate out the causative effects of advertising for each of the services, even though the results are supposed to be applicable to each service.

2. "Hypothesis 2: The observed responses are a function only of the mix in advertising expenditures."

a. The formulation used is as follows:

$$Y = a + b \ln P + cR$$

where P represents percent joint advertising. This formulation has the following deficiencies:

(1) Using the ratio of joint to total advertising as in the variable P implies that spending \$1 on joint and \$2 on total advertising has exactly the same effect as spending \$1,000,000,000 joint and \$2,000,000,000 total advertising, an absurd assumption that is built into this and the model of hypothesis 3.

(2) Since P is joint advertising divided by total advertising it follows that the latter is joint plus service advertising. Hence, irrespective of the data, the model asserts that "joint" and "joint plus service advertising" oppose one another in their effects on recruitment. This again is a result of the model and not the data.

a. Using $\ln P$ as in this model implies compression of the differences between the cells and thus moves in a direction opposite to the experiment which was justified (see above) on the need for having large differences between the cells. Note: This same comment applies to the model used for hypothesis 1.

b. Using percentages in the form of "joint" divided by "joint" plus services," biases the results in favor of joint in both hypothesis 2 and 3. For example, at the expenditure level of Mr. Carroll's example assigning \$3 to joint and \$2 to services in advertising has an effect in which a \$1 change in joint advertising has a 50% greater impact than the same \$1 change in service advertising. This impact would be reversed if the ratio of service advertising to service plus joint advertising were employed!

3. "Hypothesis 3: The observed responses are a function of the interaction between the level and mix of advertising expenditures." In this set of slides:

a. The following 3 piece function is used to represent this relation.

$$\text{(for } T \leq 5.50) \quad Y = 1.63 + .37 \ln P + 2.30 R$$

$$\text{(for } T \geq 7.01) \quad Y = .87 - .51 \ln P + 1.81 R$$

$$\text{(for } 5.50 \leq T \leq 7.01) \quad Y = 2.26 - 3.03 P^2 + 2.67 P + .57R$$

The same questions regarding the use of $\ln P$ and Y arise as in the preceding two models. In addition, the model does not reflect the verbally stated hypothesis. The formulas contain no interaction terms containing both T and P as needed to account for interactions.

b. Carroll illegitimately makes "statistical" inferences which are based on "non-statistical" choices of the segments he selected after the data were collected. ✓

c. The statistical tests are based on each of the pieces separately but then the pieces are used jointly to arrive at Carroll's recommendations. This is not a statistically valid procedure!

d. The final choices of the recommended proportions by WARC depend on the parabolic segment, which did not achieve statistical significance. Furthermore, as Mr. Carroll acknowledged in the April conference, the WARC attempts to extend this to the entire range of data also resulted in failure.

Conclusion. We have noted only some of the difficulties and deficiencies of the statistical and modeling approaches used in this WARC Study. As should already be apparent, no reliance can be placed on its results. In addition our own studies reach almost exactly opposite conclusions in all respects.

As noted above in the first part of this report, the approach used by WARC was almost certain to encounter difficulties in both design and execution. Recourse to intermittent studies of a "causal variety" needs to be supplemented by a more systematic basis for monitoring and controlling this very important class of activities.

Our recommendation is that DOD should install a continuous and systematic basis for monitoring and evaluating what each of the services are doing. The system should be capable of identifying inefficiencies, if any are present, and it should also provide a basis for allocating and reallocating budgets between the services and between the services and joint advertising as events may warrant.

Finally, it should also provide a basis for responding to congressional and other inquiries on a timely basis with up-to-date information and thereby also indicate when supplemental studies should be indicated if and when warranted.

CENTER FOR CYBERNETIC STUDIES
THE UNIVERSITY OF TEXAS AT AUSTIN

CBA3.202 • Austin, Texas 78712-1177 • (512) 471-1821

July 11, 1986

TO: Commander
U.S Army Recruiting Command
Attention: USARCPAE-ARA (Dr. Klopp)
Ft. Sheridan, IL 60037

FROM: A. Charnes *A. Charnes*
Director, Center for Cybernetic Studies

SUBJECT: Critique of Draft of Final Report of Ad-Mix Study

Attached herewith is a critique we have prepared on the "Final Report" of the Ad-Mix study sent to us by Major David Thomas of USAREC on June 18, 1986. We note that LTC Ford's memo of June 13, 1986, which accompanied this report indicates that it is not final but is rather only a "draft" and the appendices referenced in this draft "are still to come." Finally, the WCAR data as needed for computational checks and data accuracy are still not fully available to us.

Under these circumstances, it is not possible to prepare our critique in final form. Our analysis of what is now available indicates that the report as now drafted and submitted by WCAR is faulty to the point of being both seriously in error and misleading in important respects and, hence, no reliance should be placed on either its findings or its budgetary recommendations. It therefore seemed best to get our critique to you in the attached form in order to forestall possibly serious errors in policy decisions flowing from the WCAR report. The attached critique should then be regarded as extending and documenting what we said about the WCAR presentation in our memorandum of April 23, 1986.

In any event, the attached critique forms only one part of our study. Another part, based on our work using Data Envelopment Analysis, was sent to you on March 27, 1986. After you have had a chance to review these materials we shall be glad to meet with you and members of your staff to consider additional steps that might be desirable to complete or extend this work.

AC/cc

cc: W.W. Cooper ✓
B. Golany
P. Brockett

END

12-86

DTIC